

INFORMATION STUDIES DAYS 2022

Open image collections – sources of research data?

Elina Late

Tampere University

elina.late@tuni.fi

<https://orcid.org/0000-0002-3232-1365>

Sanna Kumpulainen

Tampere University

sanna.kumpulainen@tuni.fi

<https://orcid.org/0000-0002-7016-257X>

Keywords: digital images, information interaction, research data

How to cite: Late, E., & Kumpulainen, S. (2022). Open image collections – sources of research data? *Informaatiotutkimus*, 41(2–3), 88–91. <https://doi.org/10.23978/inf.122591>

This article is licensed under the terms of the CC BY-NC-SA 4.0 -license

Introduction

Images, as information sources, create an ever-increasing share of today's information horizons being one of the megatrends on the Internet (Meeker, 2019). Various image collections are open to users and different uses, one being the academic utilization as sources of research data. Openness of data in this sense extends, particularly in the field of social sciences and humanities, beyond the traditional thinking from research data repositories to the large non-academic collections of data, such as social-media images shared in Twitter or Instagram etc., to newspaper collections and curated historical image collections. These collections and services are never born or created for research purposes, but they are often used in research. However, from being seemingly open to general audience, their academic use is often very difficult or even inhibited. By now many supports available for the researchers have been designed from data-centric or information management perspective. These approaches lack of considering the researchers' perspective and understanding what the requirements are for the images when used as research data. By learning the nature of images as research data and how scholars use images as research data, we can better understand and support the needs scholars have.

ImAccess project

The ImAccess project funded by the Academy of Finland (2022–2026) investigates the use of digital images as research data and the interactions with the data. The project a) examines the information horizons of the researchers who are using open images as research data (c.f., Sonnendwald, 1999; Huvila, 2009) b) suggests intervention points based on the horizon maps and cognitive access points identified in these (see e.g. Kumpulainen et al., 2020), and c) creates and evaluates a prototype information retrieval tool with machine learning -based content analysis, which is aimed at supporting the desired access points defined in the information horizons.

The project builds on the theoretical framework of task-based information interaction to study how information is identified, selected, collected and synthesized using multiple information systems and sources (Byström & Kumpulainen, 2020; Järvelin et al., 2015; Kumpulainen & Järvelin, 2010).

Case: War time photographs

The first case study conducted in the project focuses on the use of Finnish war time photographs (FWPA). The digital collection was published online in 2013 and is open to public (<http://sa-kuva.fi/>). The photograph archive is a unique digital collection containing around 160,000 pictures from the Second World War during 1939–1945. The images portray life on the home front, events and operations at the front, the war industry, leisure time at the front, damages in bombings and the evacuation of Finnish Carelia.

Although much research has been done on contents of various collections (e.g. Late & Kumpulainen, 2022; Kumpulainen & Late, 2022), there has been little research into how people search and use historical images. Thus, The FWPA offers interesting case for the ImAccess project. For information studies perspective the case is fruitful because, searching an image from the archive may be difficult since it is relying on existing textual descriptions and metadata which may be impartial. Usage of the archive may require special skills due to the historical context of the descriptions.

A first part of the study was conducted in 2021 to get a quantitative data of the information practices of the archive users (Kumpulainen & Ruotsalainen, 2022). The study is continued by qualitative approach including interview and demonstration data of the archive users. In spring 2022 15 archive users using the collection for research and writing tasks were interviewed. During the interviews the users also demonstrated their typical uses of the collection. This rich data provides information about the uses of the archive, the searching behavior of the users, and the problems faced by the users.

Few efforts have been made towards understanding the image searchers' search goals and task-based information needs and existing research is rather old in the fields of information interaction and interactive information retrieval. Therefore, there is an urgent need for studying task-based information requirements and how to best provide task-relevant image information related the needs. In image search, the information items are visual instead of textual documents. Nevertheless, the contents are often retrieved based on textual descriptions of the images that do not necessarily match with the desired access points. The findability of the image collections could be improved by providing access points that are created with automatic means, e.g., by recognition of other objects and landscapes from the images and adding some metadata to be used in provision better visual browsing opportunities.

Acknowledgement

This research is funded by Academy of Finland grant number 351247.

References

- Byström, K., & Kumpulainen, S. (2019). Vertical and horizontal relationships amongst task-based information needs. *Information Processing & Management*, 102065. <https://doi.org/10.1016/j.IPM.2019.102065>
- Huvila, I. (2009). Analytical information horizon maps. *Library & Information Science Research*, 31(1), 18–28. <https://doi.org/10.1016/j.lisr.2008.06.005>
- Järvelin, K., Vakkari, P., Arvola, P., Baskaya, F., Järvelin, A., Kekäläinen, J. . . . Sormunen, E. (2015). Task-based information interaction evaluation: The viewpoint of program theory. *ACM Transactions on Information Systems*, 33(1), 1–30. <https://doi.org/10.1145/2699660>
- Kumpulainen, S., & Järvelin, K. (2010). Information interaction in molecular medicine: Integrated use of multiple channels. In *IIIX 2010 - Proceedings of the 2010 Information Interaction in Context Symposium* (pp. 95–104). <https://doi.org/10.1145/1840784.1840800>
- Kumpulainen, S., Keskustalo, H., Zhang, B., & Stefanidis, K. (2020). Historical reasoning in authentic research tasks: Mapping cognitive and document spaces. *Journal of the Association for Information Science and Technology*, 71(2), 230–241. <https://doi.org/10.1002/asi.24216>
- Kumpulainen, S., & Ruotsalainen, H. (2022). Searching Wartime Photograph Archive for Serious Leisure Purposes. In *Linking Theory and Practice of Digital Libraries. TPD L 2022* (pp. 81–92). Lecture Notes in Computer Science, vol 13541. Springer. https://doi.org/10.1007/978-3-031-16802-4_7
- Late, E., & Kumpulainen, S. (2021). Interacting with digitised historical newspapers: understanding the use of digital surrogates as primary sources. *Journal of Documentation*, 78(7), 106–124. <https://doi.org/10.1108/JD-04-2021-0078>
- Meeker, M. (2019). *Internet_Trends_2019_CH(1).pdf*. In *Internet trends 2019*. Bond, Code Conference 2019. <https://www.bondcap.com/report/itr19/>
- Sonnenwald, D. H. (1999). Evolving perspectives of human information behavior: Contexts, situations, social networks and information horizons. In *Exploring the contexts of information behavior: Proceedings of the Second International Conference in Information Needs*. Taylor Graham.